

Détection de l'incertitude et de la négation : un état de l'art

Clément Dalloux¹

(1) IRISA - CNRS, Campus de Beaulieu, 35042 Rennes, France

clement.dalloux@irisa.fr

RÉSUMÉ

L'un des objectifs de nos travaux, à terme, est de transformer un corpus de documents médicaux en données structurées pour en faciliter l'exploitation. Ainsi, il est nécessaire non seulement de détecter les concepts médicaux évoqués, mais aussi d'intégrer un processus capable d'identifier le contexte dans lequel est évoqué chaque concept médical. Dans cet article, nous revenons principalement sur les systèmes par apprentissage supervisé qui ont été proposés pour la détection de l'incertitude et de la négation. Ces dix dernières années, les travaux pour détecter l'incertitude et la négation dans les textes en anglais ont donné des résultats satisfaisants. Cependant, il existe encore une marge de progression non-négligeable.

ABSTRACT

Identifying uncertainty and negation's cues and scope : State of the art

One of the goals of our endeavours is to turn a corpus of medical documents into more easily readable structured data. Thus, it is necessary to add a process capable of identifying the context in which each medical concept is used. In this article, we mainly review the various systems proposed for negation and uncertainty detection that include a machine learning method. This last decade, studies undertaken in order to detect uncertainty and negation in texts have given satisfying results, however, there is still room for improvement.

MOTS-CLÉS : étiquetage automatique, incertitude, négation, apprentissage artificiel, apprentissage supervisé.

KEYWORDS: automatic labeling, uncertainty, speculation, negation, machine learning, supervised learning.

1 Introduction

La détection d'évènements incertains et niés dans les textes est une thématique de recherche à laquelle les chercheurs en TAL s'intéressent depuis une quinzaine d'années. Cela est notamment le cas dans le domaine biomédical dans lequel s'inscrit nos travaux de thèse ; en effet, beaucoup d'informations sont apportées sous forme non structurée (c'est-à-dire sous la forme de texte libre) et il est important de savoir détecter les incertitudes et les négations qui sont utilisées pour exprimer des impressions, faire des hypothèses ou bien donner des résultats négatifs. D'ailleurs, l'élaboration du corpus BioScope ainsi que la tâche commune qui eut lieu lors de la *Conference on Computational Natural Language Learning* de 2010 (Farkas *et al.*, 2010) montrent l'intérêt de cette tâche pour la fouille de textes biomédicaux (cf. infra).

Cet article présente un état de l'art des travaux sur la détection automatique de la négation et de

l'incertitude, qui constitueront le point de départ de nos travaux. Ainsi, nous présentons les approches proposées ces dernières années selon deux axes, la détection des marqueurs, c'est-à-dire des termes qui expriment une négation ou une incertitude, et la détection de la portée, c'est-à-dire des termes qui sont impactés par ces phénomènes.

L'article est structuré de la manière suivante. Dans la section qui suit, nous définissons ces phénomènes, nous donnons une liste de marqueurs typiques, et nous analysons les problèmes liés à leur détection. Dans la section 3, nous présentons les ensembles de données qui ont été annotés dans le but d'automatiser leur détection, soit pour procéder à un entraînement (apprentissage supervisé) ou bien pour extraire divers types de descripteurs. Les techniques par système expert et apprentissage artificiel dédiés à la détection des marqueurs et de la portée sont abordées dans la section 4. Enfin, nous présentons nos conclusions et perspectives dans la dernière section.

2 Détection des phénomènes : Analyse du problème

Le Dictionnaire de l'Académie française (9^{ème} édition) définit la négation comme *l'énoncé qui rejette comme faux, qui déclare faux, une proposition ou un jugement, ou bien, du point de vue de la linguistique, comme la construction utilisée pour nier ou pour exprimer le refus* et l'incertitude comme *le caractère de ce qui est incertain, douteux ou imprévisible*. Ces définitions génériques n'indiquent aucune piste pour permettre de repérer ces phénomènes dans les textes, mais laissent supposer une grande variété de réalisations linguistiques. En effet, ces constructions s'expriment par l'utilisation de marqueurs que l'on retrouve dans la plupart des catégories grammaticales. Nous présentons ci-dessous un ensemble de marqueurs de négation parmi les plus courants pour l'anglais, affixes compris, et des principaux marqueurs d'incertitude. Il s'agit bien sûr d'un ensemble non exhaustif et ouvert :

— Négation

- Auxiliaire : *Cannot*.
- Noms : *Absence, lack*.
- Verbes : *lack, exclude, fail, favor over, miss, rule out*.
- Conjonctions : *neither...nor, rather than*.
- Adjectifs : *absent, impossible, negative, unable*.
- Adverbes : *no, never, no longer, not*.
- Prépositions : *except, instead of, with the exception of, without*.
- Affixes : *un-, in-, im-, dis-, non-, -less, etc.*

— Incertitude

- Auxiliaires : *may, might, can, would, could, should*.
- Verbes : *suggest, question, presume, suspect, indicate, suppose, seem, appear, favor*.
- Conjonctions : *or, and/or, either...or*.
- Adjectifs ou adverbes : *probable, likely, possible, unsure*.

La principale difficulté pour la tâche de détection des marqueurs est d'être capable de résoudre les ambiguïtés, c'est-à-dire qu'il faut que le système soit capable de différencier les contextes où les termes sont utilisés comme marqueur de ceux où il ne le sont pas. Par exemple, l'auxiliaire *can* peut être utilisé sans marquer l'incertitude (1). Cela vaut pour la plupart des termes qui servent à exprimer l'incertitude.

1. It can only be activated in a tissue-specific manner by elements that lie downstream of the initiation site.

Dans la plupart des cas, l'effet du marqueur sur la phrase s'étend sur tout ou partie des unités lexicales. Celles impactées forment ce que l'on appelle communément **la portée** (*scope* en anglais). Identifier

automatiquement ces unités est une tâche complexe pour plusieurs raisons. Par exemple, la présence d'un ou plusieurs marqueurs de négation n'implique pas toujours que la phrase soit négative, elle peut être incertaine (1), voire positive si accompagnée par un adverbe de fréquence (2). La portée peut s'étendre des deux côtés du marqueur (3 et 4) et être discontinue (5 et 6). La méthode automatique devra donc être capable de capturer la dépendance entre le sujet et le prédicat nié, tout en excluant le syntagme verbal positif. Par ailleurs, plusieurs portées peuvent se chevaucher dans les énoncés qui contiennent plusieurs marqueurs (7 et 8). Ces portées doivent être repérées indépendamment l'une de l'autre.

1. *A small amount of adenopathy **cannot be** completely **excluded**.*
2. *Treated early, cancer is **not** always a death sentence.*
3. *Horseshoe **or** pelvic kidneys, ureteral abnormalities.*
4. *asthma **may** limit ability to participate in the exercise intervention.*
5. *The patient visited the hospital but did **not** meet the doctor.*
6. *The patient is coming to the hospital but **might** be late.*
7. *The patient did **not** take his medication because he was not feeling sick.*
8. *The patient did not take his medication because he was **not** feeling sick.*

Les publications liées à ces tâches de détection ou de classification ont principalement été publiées dans le cadre de la description du corpus BioScope (Vincze *et al.*, 2008), ainsi que lors des tâches communes de CoNLL-2010 (Farkas *et al.*, 2010) portant sur la détection de l'incertitude, et SEM-2012 (Morante & Blanco, 2012) portant sur la détection de la portée et de la cible de la négation. Nous présentons BioScope ainsi que les ensembles de données utilisés dans la section suivante.

3 Les données

Ces dernières années, avec la démocratisation des techniques d'apprentissage supervisé, de nombreux corpus de spécialité ont été annotés afin d'entraîner des modèles pour la détection automatique d'événements et de phénomènes linguistiques. Cette section est dédiée à la description des données annotées pour l'incertitude et/ou la négation.

3.1 BioScope

Le corpus BioScope (Vincze *et al.*, 2008), est constitué de 3 sources textuelles différentes, à savoir de comptes-rendus d'examens radiologiques, d'articles de FlyBase et BMC Bioinformatics, ainsi que de résumés d'articles scientifiques issus du corpus GENIA¹. Selon les consignes d'annotations² préétablies, cet ensemble de données est annoté au niveau des marqueurs de l'incertitude et de la négation, et au niveau de la phrase pour marquer leur portée linguistique. Les portées discontinues ne sont pas prises en compte. Le tableau 1 quantifie la composition du corpus, avec P_{neg} phrases négatives, M_{neg} marqueurs de la négation, P_{inc} phrases incertaines, M_{inc} marqueurs de l'incertitude.

Illustrées en figure 1, les données sont disponibles au format XML, où chaque phrase et chaque paire portée/marqueur (*xcope* et *cue*) sont indexées par un identifiant unique.

1. <http://www.geniaproject.org/>

2. <http://bit.ly/2mDnAoL>

	Compte-rendus	Articles	Résumés
Documents	1 954	9	1 273
Phrases	6 383	2 670	11 871
P_{neg}	13,55%	12,70%	13,45%
M_{neg}	877	389	1 848
P_{inc}	13,39%	19,44%	17,70%
M_{inc}	1 189	714	2 769

TABLEAU 1 – Statistiques du corpus BioScope

<sentence id="S1.7">These results <xcope id="X1.7.2"><cue type="speculation" ref="X1.7.2">indicate that</cue>
in monocytic cell lineage, HIV-1 <xcope id="X1.7.1"><cue type="speculation" ref="X1.7.1">could</cue> mimic
some differentiation/activation stimuli allowing nuclear NF-KB expression</xcope></xcope>.</sentence>

FIGURE 1 – Extrait du corpus BioScope

3.2 CoNLL-2010

Dans le cadre des tâches communes (shared task) de la *Conference on Computational Natural Language Learning* de 2010 (Farkas *et al.*, 2010), trois jeux de données ont été mis à la disposition des participants. La première tâche consistait à identifier les phrases qui contiennent des informations incertaines ou douteuses à l’aide de textes où seuls les marqueurs sont annotés. Le premier jeu de données est composé de paragraphes issus de Wikipedia où apparaissent des « weasel words », c’est-à-dire, des mots et phrases qui donnent l’impression qu’une déclaration spécifique ou significative a été faite, alors qu’il s’agit en réalité d’une revendication vague et ambiguë. Le deuxième jeu de données est issu de BioScope, mais sans les compte-rendus cliniques. La seconde tâche consistait à identifier la portée de l’incertitude dans les données biomédicales, les marqueurs d’incertitude et leurs portées y sont annotés.

3.3 *SEM-2012

Prononcé « *starsem* » (*Conference on Lexical and Computational Semantics : *SEM*) propose chaque année une tâche commune. Celle de 2012 s’est attaquée à deux aspects de la négation : la détection de la portée et de la focalisation qui correspond à la partie de la portée la plus explicitement niée. Dans l’exemple suivant, la portée est entre crochets, l’objet entre * et la focalisation soulignée.

[John had] never [*said* as much before].

L’ensemble des données est composé d’un roman et de trois nouvelles de Sherlock Holmes par Conan Doyle, *The Hound of the Baskervilles* pour l’entraînement, *The Adventures of Wisteria Lodge* pour le développement, et *The Adventure of the Red Circle* et *The Adventure of the Cardboard Box* pour les tests. Toutes les occurrences de la négation sont annotées, ce qui donne 1 056 phrases sur 3 899. Pour chaque occurrence, le marqueur et la portée sont annotés, ainsi que l’objet de la négation, s’il y en a un. Les marqueurs et les portées peuvent être discontinus dans ce corpus. Le guide d’annotation a été publié également (Morante *et al.*, 2011). Dans le tableau 2, nous présentons un exemple d’annotation, avec #C numéro de chapitre, #P numéro de phrase, #U position de l’unité lexicale dans la phrase.

Partant du constat que les descripteurs lexicaux et syntaxiques (*lemme*, *POS*, *Chunk*) ont été massivement utilisés dans le cadre de CoNLL-2010, les organisateurs de *SEM-2012 les ont inclus d’office

#C	#P	#U	Unité	Lemme	POS	Chunk	Marqueur	Portée	Évènement
baskervilles01	150	0	“	“	“	(SQ*	—	—	—
baskervilles01	150	1	Then	Then	RB	(ADVP*)	—	—	—
baskervilles01	150	2	had	have	VBD	*	—	had	—
baskervilles01	150	3	you	you	PRP	(NP*)	—	you	—
baskervilles01	150	4	not	not	RB	(ADVP*	not	—	—
baskervilles01	150	5	better	well	RB	*)	—	better	—
baskervilles01	150	6	consult	consult	VB	(VP*	—	consult	—
baskervilles01	150	7	him	him	PRP	(NP**))	—	him	—
baskervilles01	150	8	?	?	.	*	—	—	—
baskervilles01	150	9	”	”	”	*)	—	—	—

TABLEAU 2 – Extrait – Données d’entraînement *SEM-2012

dans leur corpus. En outre, ces données annotées peuvent aussi aider à la conception de systèmes experts spécifiquement conçus pour l’ensemble de données en question. Dans la section suivante, nous attaquons la description des travaux de recherche sur la détection automatique antérieurs et postérieurs à la mise à disposition de ces données.

4 Détection automatique : description des approches

D’une manière générale, il existe deux familles d’approche pour aborder la détection automatique des marqueurs : l’utilisation de systèmes experts, raisonnant à partir de faits et règles connus pour répondre à des questions précises, et la classification par apprentissage automatique, à l’aide de méthodes telles que les champs aléatoires conditionnels (*Conditional Random Fields* ou CRFs), les machines à vecteurs support (SVM) ou les réseaux de neurones.

4.1 Système expert

Un système expert est un programme qui raisonne à partir d’informations symboliques et utilise des règles expertes afin d’essayer d’atteindre les performances du niveau d’un expert. Pour des tâches spécifiques, il est possible de construire un système qui raisonne aussi bien que les spécialistes de ces domaines. Partant de cette hypothèse, les premiers systèmes dédiés à la détection de la négation sont proposés sous cette forme. Ainsi, au début des années 2000, Chapman *et al.* (2001) propose l’algorithme *NegEx* qui utilise les expressions régulières pour détecter les phrases négatives et identifier les termes médicaux à l’intérieur de la portée de la négation. La même année, Mutalik *et al.* (2001) propose *Negfinder*. Le système combine un analyseur lexical (lexeur) qui utilise des expressions régulières afin de générer un automate fini et un analyseur syntaxique (parser) qui repose sur un sous-ensemble restreint de la grammaire non contextuelle LALR³. Ainsi, *NegFinder* permet d’identifier les concepts impactés par la négation dans les textes médicaux lorsqu’ils sont proches du lemme marquant la négation. Une extension des travaux précédents est développée par (Elkin *et al.*, 2005). Le système utilise le *Mayo Vocabulary Server* afin de récupérer un ensemble de concepts médicaux présents dans les fragments de phrase. Ensuite, des règles expertes assignent à chaque concept un descripteur qui indique si le concept est une déclaration positive, négative ou incertaine.

3. Look-Ahead Left Recursive

Par la suite, les systèmes experts ont majoritairement impliqué les traits syntaxiques. Huang & Lowe (2007) proposent d’automatiser la détection de la négation en combinant les expressions régulières et l’analyse grammaticale. Les négations sont classées sur la base des catégories syntaxiques et sont indiquées dans des arbres syntaxiques. Cette approche hybride permet d’identifier les concepts marqués par la négation dans les rapports de radiologie même s’ils sont situés à distance des termes négatifs. Özgür & Radev (2009) émettent l’hypothèse que l’on peut caractériser la portée d’un marqueur de l’incertitude par sa catégorie grammaticale et la structure syntaxique de la phrase. Ils développent des règles pour détecter la portée d’un marqueur à partir des noeuds dans un arbre syntaxique. Øvrelid *et al.* (2010) construisent un petit ensemble de règles qui définissent la portée de chaque marqueur. Pour développer ces règles, les auteurs exploitent les informations contenues dans le guide d’annotation du corpus BioScope, ainsi qu’une analyse manuelle des données d’entraînement pour trouver les interactions des constructions pour divers types de marqueurs. Kilicoglu & Bergler (2010) proposent une méthodologie pour la détection de l’incertitude essentiellement basée sur des règles, qui repose sur une combinaison d’informations lexicales et syntaxiques. L’information lexicale est représentée par un simple dictionnaire, alors que l’information syntaxique nécessaire est identifiée grâce aux arbres d’analyse de constituants et aux arbres d’analyse des dépendances retournés par le Stanford Parser⁴. Apostolova *et al.* (2011) présentent ScopeFinder, un système expert pour la détection de la portée de la négation et de l’incertitude, dont les règles sont construites à partir des patrons lexico-syntaxiques extraits automatiquement du corpus BioScope. La méthode de construction du système fait qu’il est adaptable aux données de différents domaines.

Les systèmes experts présentés dans cette section arrivent à obtenir de bons résultats sur les données biomédicales pour lesquelles ils sont développés. Cependant, pour être appliqués efficacement, ces systèmes doivent être spécifiquement adaptés aux langues et types de données concernés. Les systèmes par apprentissage artificiel souffrent moins de ce problème.

4.2 Apprentissage artificiel

Depuis plusieurs années, les méthodes d’apprentissage artificiel permettant de manipuler de grands volumes de données se sont popularisées en TAL. Pour les tâches de classification que nous étudions, l’apprentissage supervisé est principalement utilisé puisque plusieurs ensembles de données annotées sont disponibles. Les systèmes les plus performants utilisent principalement les SVM, les CRF, ainsi que les réseaux de neurones. D’ailleurs, il nous semble que le premier système pour la détection de l’incertitude exploite l’apprentissage artificiel avec un SVM (Light *et al.*, 2004). Dans ce travail, les auteurs utilisent le système pour sélectionner les phrases spéculatives dans des résumés d’articles de MEDLINE pour démontrer que la détection de l’incertitude peut être efficacement traitée en se basant sur les lemmes qui l’impliquent. Dans la plupart des travaux, les systèmes combinent un module de détection des marqueurs de négation, et un module de détection de la portée de cette négation. Nous les examinons ci-dessous.

4.2.1 Détection des marqueurs

Dans cette section, nous présentons les systèmes par apprentissage supervisé les plus performants pour la détection des marqueurs de l’incertitude et de la négation. Dans le cadre des tâches communes de CoNLL-2010, les systèmes utilisent, pour la plupart, le GENIA Tagger afin d’obtenir les lemmes,

4. <https://nlp.stanford.edu/software/lex-parser.shtml>

l'étiquetage morpho-syntaxique (POS), ainsi que l'analyse syntaxique de surface (*Shallow parsing, chunking*). D'autre part, bien que la tâche commune de *SEM-2012 ne porte pas sur la détection des marqueurs, les systèmes soumis reposent tous sur un sous-système de détection des marqueurs dont la description et les résultats sont donnés. Les résultats sont évalués à l'aide de trois métriques : la précision, qui quantifie la pertinence de l'étiquetage, le rappel, qui quantifie la sensibilité de l'étiquetage, ainsi que la moyenne harmonique de la précision et du rappel noté F_1 . Nous présentons les résultats des systèmes les plus performants dans le tableau 3 en fin de section.

Le système proposé par Tang *et al.* (2010) est le plus performant sur les deux ensembles de données de CoNLL. Le système classe les unités selon une représentation en BIO (*Begin, Inside, Outside*). La détection des marqueurs se fait en trois étapes. La première consiste à pré-traiter les données avec le GENIA Tagger. La seconde opère deux classifications indépendantes et adaptées à l'étiquetage de séquences (CRF++-0.53⁵ et SVM^{hmm}⁶) sur les données pré-traitées. La dernière étape consiste à entraîner un CRF en utilisant les prédictions de l'étape précédente. Les descripteurs utilisés sont : les lemmes et les séquences d'étiquetage morpho-syntaxique de l'incertitude retournés par chaque classifieur, la classification des unités en B, I ou O dont l'étiquetage est identique pour les deux classifieurs, la position de l'unité si elle est la dernière unité du marqueur prédite par chaque classifieur, ainsi que l'ensemble des descripteurs utilisés par l'étape précédente.

Le système de Li *et al.* (2010b), deuxième sur le corpus de Wikipedia et troisième sur les données biomédicales, classe les unités selon une représentation en BIO. Le classifieur est entraîné en utilisant un CRF (CRF++ 0.51) avec les paramètres par défaut. Ils utilisent un algorithme glouton pour sélectionner un meilleur ensemble de descripteurs (unité, lemme, POS, chunk) pour le classifieur selon les résultats obtenus sur les données de développement. En post-traitement, ils utilisent des règles expertes pour classer correctement certains marqueurs oubliés ou incorrectement classés.

Le système de Zhou *et al.* (2010), second sur le corpus de données biomédicales, identifie les marqueurs candidats à partir d'une liste de mots clés puis applique un CRF pour corriger les erreurs d'identification. Les marqueurs récupérés sont utilisés comme descripteurs pour la détection par CRF. Les lemmes sont remplacés par les radicaux.

Extension de leurs travaux précédents, le système proposé par Velldal *et al.* (2012) traite l'ensemble de marqueurs en tant que classe fermée. Ce faisant, ils expliquent que l'on peut réduire considérablement le nombre d'exemples et conséquemment le nombre de descripteurs présentés au classifieur sans perdre en rappel. Le système utilise un SVM appliqué en utilisant de simples descripteurs n-gramme sur les mots et lemmes, à la droite et la gauche des marqueurs candidats.

Read *et al.* (2012) et Lapponi *et al.* (2012) sont deux systèmes proposés par des chercheurs de l'université d'Oslo dans le cadre de *SEM-2012. Leur système de détection des marqueurs est une adaptation de leur précédent système de classification (Velldal *et al.*, 2012). La principale modification est la prise en charge de l'identification des marqueurs de négation affiaux annotés dans les données d'entraînement et de développement. Le SVM utilise des descripteurs qui ciblent spécifiquement les marqueurs affiaux. Le premier descripteur enregistre des n-grammes du début et de la fin de la base à laquelle un affixe s'accroche. Par exemple, dans le cas du marqueur *impossible* les n-grammes {possi, poss, pos, ...} et {sible, ible, ble, ...} sont enregistrés et sont joints avec l'affixe *im* et l'étiquette POS *JJ*. Le second descripteur cherche à imiter l'effet d'une recherche dans un lexique pour les séquences restantes auquel un affixe s'attache aussi, en vérifiant son statut en tant que forme de base

5. <https://taku910.github.io/crfpp/>

6. https://www.cs.cornell.edu/people/tj/svm_light/svm_hmm.html

indépendante et son étiquetage POS. La raison derrière ce descripteur est que l'occurrence d'une séquence telle que *un* dans un mot comme *underlying* doit être considérée comme peu probable d'être un marqueur étant donné que la première partie de la séquence restante constitue très rarement le début d'un mot. Les marqueurs multi-mots tels que *neither...nor* ou bien l'expression *on the contrary* sont trop rares pour être détectés par le classifieur. Par conséquent, un petit ensemble de règles expertes est développé afin de les détecter en post-traitement.

Après avoir écarté les affixes marquant la négation, le système de Chowdhury (2012) collecte automatiquement un vocabulaire de toutes les unités lexicales qui ne sont pas des marqueurs de négation et dont la taille dépasse trois caractères dans les données d'entraînement. Ensuite, le système les utilise pour extraire des descripteurs qui pourraient être utiles pour identifier des affixes marqueurs de négation potentiels (**un*able*). Une liste d'expressions exprimant très probablement la négation est créée à partir des données d'entraînement en se basant sur leurs fréquences. La liste contient les termes suivants : *nor, neither, without, nobody, none, nothing, never, not, no, nowhere, non*. Les marqueurs de négation affixaux sont identifiés si l'unité est prédite en tant que marqueur par le classifieur et possède l'un des affixes suivants qui sont collectés à partir des données d'entraînement : *less, un, dis, im, in, non, ir*. Les lemmes sont passés en minuscule dans l'ensemble descripteurs. Un post-traitement supplémentaire est effectué pour annoter certains marqueurs présents dans les données d'entraînement mais oubliés par le classifieur lors de la prédiction sur les données de développement. Ces expressions sont *neither, nobody, save for, save upon, et by no means*. Une vérification spéciale est faite pour l'expression *none the less* qui est marquée comme une non-négation dans les données d'entraînement. Finalement, un modèle de CRF est entraîné sur les descripteurs collectés et utilisé pour prédire les marqueurs de négation sur les données de test.

Données	Système	Apprentissage	Précision	Rappel	F_1
Wikipedia	Tang <i>et al.</i> (2010)	CRFs+SVM	63,0	25,7	36,5
	Li <i>et al.</i> (2010b)	CRF	76,1	21,6	33,7
Biomédical	Tang <i>et al.</i> (2010)	CRFs+SVM	81,7	81,0	81,3
	Zhou <i>et al.</i> (2010)	CRF	83,1	78,8	80,9
	Velldal <i>et al.</i> (2012)	SVM	84,8	77,2	80,8
	Li <i>et al.</i> (2010b)	CRF	87,4	73,4	79,8
*SEM-2012	Read <i>et al.</i> (2012)	SVM	91,42	92,80	92,10
	Lapponi <i>et al.</i> (2012)	SVM	89,17	93,56	91,31
	Chowdhury (2012)	CRF	93,41	91,29	92,34

TABLEAU 3 – Résultats pour la détection des marqueurs pour l'incertitude (CoNLL) et la négation (*SEM-2012) (scores en pourcentage, les meilleurs scores sont en gras)

Pour la détection des marqueurs d'incertitude comme de négation, les différences de performance entre les systèmes les plus efficaces sont ténues. Ils sont même équivalents si l'on prend en considération les légères variations de scores qui peuvent intervenir selon les entraînements. Comme souvent, une haute précision entraîne une chute du rappel et inversement. Les systèmes exploitent de nombreux descripteurs et reposent donc sur les performances des étiqueteurs qui restent imparfaites. En outre, la plupart des auteurs cités notent la sous-représentation de certains marqueurs dans les données d'entraînement rend nécessaire l'application de règles de post-traitement dépendantes de la langue.

4.2.2 Détection de la portée

Dans cette section, nous présentons les systèmes par apprentissage supervisé les plus performants pour la détection de la portée de l'incertitude et de la négation. Les résultats obtenus sur le corpus BioScope sont évalués en pourcentage de portées correctement identifiées. Pour *SEM-2012, les résultats sont évalués au niveau des unités lexicales identifiées et des portées correctement identifiées (précision, rappel, F_1). Nous présentons les résultats des systèmes les plus performants dans les tableaux 4 et 5 en fin de section.

Li *et al.* (2010a) proposent une approche pour l'apprentissage de la portée de la négation utilisant un système d'analyse sémantique superficielle simplifiée qui combine des règles expertes, l'élagage et la classification binaire par SVM. Étant donné un arbre syntaxique et un prédicat lui appartenant, l'analyse sémantique superficielle reconnaît et assigne à tous les constituants dans la phrase le rôle sémantique leur correspondant. Le marqueur de négation peut être considéré comme le prédicat, alors que les constituants de la portée peuvent être considérés comme les arguments du marqueur de la négation. Deux règles expertes sont adoptées pour assigner le rôle d'argument de la portée sur plusieurs constituants dans un arbre syntaxique donné. Excepté le marqueur de négation et ses ancêtres, tous les constituants dans l'arbre syntaxique dont le parent couvre le marqueur sont collectés comme arguments candidats. Finalement, un classifieur binaire⁷ est appliqué pour déterminer si les arguments candidats sont valides.

Le système proposé par Velldal *et al.* (2012) combine deux approches. L'une utilise des règles expertes opérant sur des arbres syntaxiques de dépendance. La seconde utilise un SVM d'ordonnement (*ranking*) qui opère sur des noeuds d'arbres syntaxiques des constituants syntagmatiques.

L'année dernière, un système basé sur un réseau de neurones convolutionnel (CNN) a été proposé par Qian *et al.* (2016). Les descripteurs pertinents sont récupérés à partir des chemins syntaxiques entre les marqueurs et les unités candidates dans les arbres syntaxiques de dépendances et les arbres syntaxiques de constituants syntagmatiques. La couche convolutionnelle concatène ces descripteurs avec leurs positions relatives en un seul vecteur descripteur qui alimente une couche *softmax*. Cette couche cachée calcule les scores de confiance de l'étiquetage pour une classification en OBA où sont classés : O les unités lexicales ne faisant pas parties de la portée, B les unités lexicales de la portée qui apparaissent avant le marqueur, et A les unités lexicales de la portée qui apparaissent après le marqueur.

Dans la continuité de Velldal *et al.* (2012), Read *et al.* (2012) combine un système expert et l'ordonnement par SVM. Les règles expertes y sont formulées en fonction de patrons cohérents relevés en alignant constituants syntaxiques et les portées. Les règles s'appuient sur la fréquence des chemins allant du marqueur jusqu'au constituant aligné avec la portée sur la base des annotations dans les données d'entraînement (SEM-2012). L'approche motivée par les données pour la résolution de la portée implique l'apprentissage d'une fonction de rang sur les constituants syntaxiques candidats. Cela commence par la sélection de négations pour lesquelles l'arbre syntaxique contient un constituant qui s'aligne avec la portée. Ensuite, la sélection d'un candidat initial est effectuée en prenant le constituant le plus petit qui s'étend sur tous les mots dans la portée. Finalement, la génération d'un candidat subséquent est effectuée en traversant le chemin vers la racine de l'arbre. Les candidats dont la projection correspond à la portée sont étiquetés comme corrects ; tous les autres sont étiquetés comme incorrects. Expérimentant avec divers types de descripteurs, ils utilisent la fonction d'ordonnement ordinal présente dans le SVM-light toolkit pour apprendre une fonction

7. <http://svmlight.joachims.org/>

score linéaire pour choisir les portées candidates correctes.

Le système de détection de la portée de (Lapponi *et al.*, 2012) utilise l'implémentation CRF Wapiti⁸ avec une version spéciale de la séquence d'étiquetage IOB, où les marqueurs de négation sont étiquetés comme démarreurs de séquence et les unités de la portée comme constituants phrastiques (*chunk*). En plus des descripteurs fournis pour chaque unité dans le corpus (lemme, POS, constituant phrastique), étiquettes, la distance entre chaque unité et le marqueur le plus proche à droite et à gauche est extraite. Les descripteurs ont été étendus pour inclure des bigrammes et trigrammes vers l'avant et l'arrière des unités et au niveau des étiquettes POS, ainsi que les unigrammes et bigrammes de POS lexicalisés (couple forme/POS). Les représentations syntaxiques sont converties en représentations des dépendances en utilisant le Stanford Parser. Les descripteurs récupérés à partir des graphes de dépendances ont pour but de modéliser les relations syntaxiques entre chaque unité et le marqueur le plus proche. Finalement, un algorithme basé sur des règles expertes est appliqué en sortie du CRF dans le but d'assigner chaque unité dans la portée à son marqueur et déterminer les chevauchements de portées.

La méthode proposée par (Packard *et al.*, 2014) repose sur deux approches. La première est basée sur un indexeur par *minimal Recursion semantics* (MRS). La MRS est un cadre applicatif pour la sémantique computationnelle. Il peut être implémenté selon des formalismes tels que la grammaire syntagmatique guidée par les têtes et la grammaire lexicale-fonctionnelle. La seconde approche reprend le système de (Read *et al.*, 2012) et intervient pour la détection des marqueurs et pour la détection de la portée dans les cas où le *LinGo English Ressource Grammar* (ERG) ne peut pas créer de représentation sémantique fiable pour une phrase. La MRS rend explicite les relations prédicat-argument, ainsi que des informations partielles concernant la portée. Un système expert traverse les structures MRS des phrases négatives.

Fancellu *et al.* (2016) proposent deux approches par réseau de neurones pour résoudre le problème de détection de la portée de la négation. La première utilise un réseau de neurones *Feed-forward*, c'est-à-dire, un réseau de neurones artificiel où les connexions entre les unités ne forment pas de boucle. Dans ce type de réseau, l'information se déplace uniquement dans un sens, à partir des noeuds de la couche entrée, à travers les noeuds de la couche cachée et finalement des noeuds de sortie. La seconde approche, qui est la plus efficace, utilise un réseau de neurones Long Short-Term Memory (LSTM) bidirectionnel. Un LSTM bidirectionnel est la combinaison de deux structures de réseau de neurones récurrents (RNN). Le RNN bidirectionnel proposé par (Schuster & Paliwal, 1997), et le LSTM proposé par (Hochreiter & Schmidhuber, 1997). Ce réseau est motivé par le fait qu'un être humain raisonne en s'appuyant sur les connaissances qu'il a acquises et qui restent dans la mémoire. Si les réseaux de neurones « classiques » ne sont pas capables de copier ce processus, les RNN répondent à ce problème. En effet, ce sont des réseaux qui bouclent, ce qui permet aux informations de persister. Parmi les RNN, les réseaux LSTM sont les plus efficaces pour apprendre des dépendances de long-terme et sont donc plus à même de résoudre, par exemple, le problème de la portée discontinue. Le LSTM de (Fancellu *et al.*, 2016) bidirectionnel est un modèle séquentiel qui opère dans le sens de lecture et dans le sens contraire. La passe arrière est particulièrement importante dans le cas de la détection de la portée puisque les unités affectées peuvent se trouver avant le marqueur. Par ailleurs, les cellules LSTM sont plus efficaces pour retenir les informations utiles lors de la rétropropagation du gradient, qui permet de corriger les différences entre les prédictions sortantes et celles désirées en calculant le gradient de l'erreur pour chaque neurone, de la dernière couche vers la première. Le système ne s'entraîne que sur les phrases porteuses d'une négation. Le système de base prend

8. <https://wapiti.limsi.fr/>

	Système	Apprentissage	Résumé	Clinique	Article
Incertitude	Velldal <i>et al.</i> (2012)	SVM	79,56	78,69	75,15
	Qian <i>et al.</i> (2016)	CNN	85,75	73,92	59,82
Négation	Li <i>et al.</i> (2010a)	SVM	81,84	89,79	64,02
	Velldal <i>et al.</i> (2012)	SVM	74,35	90,74	70,21
	Qian <i>et al.</i> (2016)	CNN	77,14	89,66	55,32

TABLEAU 4 – Comparaison des systèmes les plus performants sur le corpus BioScope ; les résultats sont donnés en pourcentage de portées correctement identifiées (les meilleurs scores sont en gras)

	Système	Apprentissage	Précision	Rappel	F_1
ULI	Read <i>et al.</i> (2012)	SVM	81,99	88,81	85,26
	Lapponi <i>et al.</i> (2012)	CRF	86,03	81,55	83,73
	Packard <i>et al.</i> (2014)	SVM	86,1	90,4	88,2
	(Fancellu <i>et al.</i> , 2016)	BiLSTM	92,62	85,13	88,72
PCI	Read <i>et al.</i> (2012)	SVM	87,43	61,45	72,17
	Lapponi <i>et al.</i> (2012)	CRF	85,71	62,65	72,39
	Packard <i>et al.</i> (2014)	SVM	98,8	65,5	78,7
	Fancellu <i>et al.</i> (2016)	BiLSTM	99,40	63,87	77,77

TABLEAU 5 – Comparaison des systèmes les plus performants sur les données de test de *SEM-2012, évalués au niveau des unités lexicales identifiées (ULI) et des portées correctement identifiées (PCI) ; les résultats sont donnés en pourcentage (les meilleurs scores sont en gras)

en entrée une instance $I(n, c)$ où chaque mot est représenté par un vecteur n (*word-embedding*) et par un vecteur c qui détermine si le mot fait parti d'un marqueur (*cue-embedding*). Dans le but d'affiner la prédiction, le système peut aussi utiliser une représentation vectorielle de l'étiquetage morpho-syntaxique pour chaque mot.

Tous les systèmes de détection de la portée par CRF et SVM que nous avons présentés sont dépendants des performances de leurs systèmes de détection des marqueurs. La multiplication des descripteurs de plus en plus complexes rendent les systèmes difficiles à adapter pour une autre langue. Les descripteurs complexes et dépendants de la langues tels que la MRS (Packard *et al.*, 2014) ou encore les chemins syntaxiques (Qian *et al.*, 2016) dépendent d'outils externes dont les performances sont imparfaites. Le système de Fancellu *et al.* (2016) nous semble être le plus prometteur car il est facilement portable d'un langage à l'autre du moment que des données annotées sont disponibles.

4.2.3 Expériences & résultats

Le code de Fancellu *et al.* (2016) étant disponible en ligne⁹, nous avons pu mettre à l'épreuve son système et procéder à quelques changements. Nous avons d'abord mis à jour le code vers la version 1.0 de Tensorflow¹⁰, puis nous l'avons exécuté sur notre système. Bien que nous nous attendions à de légères différences entre les entraînements, nous constatons une instabilité dans les résultats obtenus. Les scores F_1 présentés dans le tableau 6 illustrent ces variations, qui atteignent jusqu'à 6 points d'écart. Il est fort probable que le faible volume de données d'entraînement (848 phrases

9. <https://github.com/ffancellu/NegNN>

10. <https://www.tensorflow.org/>

Apprentissage	Unités de la portée			Portées exactes		
	F_1min	F_1max	F_1moy	F_1min	F_1max	F_1moy
BiLSTM	88,20	89,68	88,83	72,46	78,34	76,00
BiGRU	87,28	89,58	88,39	73,08	77,49	74,87

TABLEAU 6 – Variation des scores F_1 du BiLSTM de (Fancellu *et al.*, 2016) et de notre BiGRU (les scores sont en pourcentage, les meilleurs scores sont en gras)

négatives) cause cette instabilité. D’autre part, nous avons remplacé les cellules LSTM du réseau par des cellules GRU (Gated Recurrent Unit – Cho *et al.* (2014)) qui, en fusionnant plusieurs étapes du LSTM, permettent un calcul simplifié et plus rapide (40% de gains sur notre système). L’ensemble de notre code est accessible sur <https://github.com/Kureman/NegNN>.

5 Conclusions

Entre 2008 et 2012, les travaux sur la détection automatique de la négation et de l’incertitude en langue anglaise se sont multipliés grâce à la sortie de BioScope ainsi que les tâches communes de CoNLL-2010 et *SEM-2012. Depuis, plusieurs travaux innovant, principalement axés sur la détection de la portée, ont vu le jour.

Dans cet article, nous avons d’abord présenté les marqueurs ainsi que les différents types de portées posant problème et que l’on peut rencontrer dans les données dont nous disposons. Nous avons ensuite présenté les corpus annotés dont les chercheurs ont disposé pour entraîner des modèles par apprentissage automatique, pour extraire des descripteurs supplémentaires, ou bien pour la conception de systèmes experts. La revue de ces différents travaux nous a montré que ces systèmes experts sont désormais délaissés au profit des approches par apprentissage artificiel, les plus performants dans les évaluations dont nous avons rendu compte. C’est donc sur ces derniers systèmes que nous nous sommes penchés, en tentant notamment de reproduire les résultats de Fancellu *et al.* (2016); nous avons remarqué une instabilité du système, due au manque de données, qui souligne une limite de ce type d’approche. Nous avons aussi proposé une modification de son approche pour obtenir des résultats équivalents mais avec un temps de calcul bien plus court.

Les travaux présentés dans cet article constituent un point de départ pour nos prochains travaux qui nécessiteront l’annotation de données en français. Nous pourrions notamment nous inspirer des consignes d’annotation disponibles pour l’anglais et du volume de données disponible dans le but d’obtenir un ensemble de données suffisamment large et varié pour stabiliser un modèle.

Remerciements

Ce travail a bénéficié d’une aide de l’État attribuée au labex COMIN LABS et gérée par l’Agence Nationale de la Recherche au titre du programme « Investissements d’avenir » portant la référence ANR-10-LABX-07-01.

Références

- APOSTOLOVA E., TOMURO N. & DEMNER-FUSHMAN D. (2011). *Automatic extraction of lexico-syntactic patterns for detection of negation and speculation scopes*, In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : short papers-Volume 2*. Association for Computational Linguistics.
- CHAPMAN W. W., BRIDEWELL W., HANBURY P., COOPER G. F. & BUCHANAN B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, **34**(5).
- CHO K., VAN MERRIËNBOER B., GULCEHRE C., BAHDANAU D., BOUGARES F., SCHWENK H. & BENGIO Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv :1406.1078*.
- CHOWDHURY M. F. M. (2012). *FBK : Exploiting phrasal and contextual clues for negation scope detection*, In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1 : Proceedings of the main conference and the shared task, and Volume 2 : Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics.
- ELKIN P. L., BROWN S. H., BAUER B. A., HUSSER C. S., CARRUTH W., BERGSTROM L. R. & WAHNER-ROEDLER D. L. (2005). A controlled trial of automated classification of negation from clinical notes. *BMC medical informatics and decision making*, **5**.
- FANCELLU F., LOPEZ A. & WEBBER B. (2016). *Neural networks for negation scope detection*, In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 1.
- FARKAS R., VINCZE V., MÓRA G., CSIRIK J. & SZARVAS G. (2010). *The CoNLL-2010 shared task : learning to detect hedges and their scope in natural language text*, In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*. Association for Computational Linguistics.
- HOCHREITER S. & SCHMIDHUBER J. (1997). Long short-term memory. *Neural Comput.*, **9**(8).
- HUANG Y. & LOWE H. J. (2007). A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American Medical Informatics Association : JAMIA*, **14**(3).
- KILICOGLU H. & BERGLER S. (2010). *A High-Precision Approach to Detecting Hedges and their Scopes.*, In *CoNLL Shared Task*.
- LAPPONI E., VELLDAL E., ØVRELID L. & READ J. (2012). *Uio 2 : sequence-labeling negation using dependency features*, In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1 : Proceedings of the main conference and the shared task, and Volume 2 : Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics.
- LI J., ZHOU G., WANG H. & ZHU Q. (2010a). *Learning the scope of negation via shallow semantic parsing*, In *Proceedings of the 23rd International Conference on Computational Linguistics*, p. 671–679. Association for Computational Linguistics.
- LI X., SHEN J., GAO X. & WANG X. (2010b). *Exploiting rich features for detecting hedges and their scope*, In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*. Association for Computational Linguistics.
- LIGHT M., QIU X. Y. & SRINIVASAN P. (2004). *The language of bioscience : Facts, speculations, and statements in between*, In *Proceedings of BioLink 2004 workshop on linking biological literature, ontologies and databases : tools for users*. Association for Computational Linguistics.

- MORANTE R. & BLANCO E. (2012). * *SEM 2012 shared task : Resolving the scope and focus of negation*, In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1 : Proceedings of the main conference and the shared task, and Volume 2 : Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics.
- MORANTE R., SCHRAUWEN S. & DAELEMANS W. (2011). Annotation of negation cues and their scope. guidelines v1.0.
- MUTALIK P. G., DESHPANDE A. & NADKARNI P. M. (2001). Use of general-purpose negation detection to augment concept indexing of medical documents : a quantitative study using the umls. *Journal of the American Medical Informatics Association : JAMIA*, **8**(6).
- PACKARD W., BENDER E. M., READ J., OEPEN S. & DRIDAN R. (2014). *Simple Negation Scope Resolution through Deep Parsing : A Semantic Solution to a Semantic Problem.*, In *ACL (1)*.
- QIAN Z., LI P., ZHU Q., ZHOU G., LUO Z. & LUO W. (2016). : Association for Computational Linguistics.
- READ J., VELLDAL E., ØVRELID L. & OEPEN S. (2012). *Uio 1 : Constituent-based discriminative ranking for negation resolution*, In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1 : Proceedings of the main conference and the shared task, and Volume 2 : Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics.
- SCHUSTER M. & PALIWAL K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, **45**(11).
- TANG B., WANG X., WANG X., YUAN B. & FAN S. (2010). *A cascade method for detecting hedges and their scope in natural language text*, In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*. Association for Computational Linguistics.
- VELLDAL E., ØVRELID L., READ J. & OEPEN S. (2012). Speculation and negation : Rules, rankers, and the role of syntax. *Computational Linguistics*, **38**(2).
- VINCZE V., SZARVAS G., FARKAS R., MÓRA G. & CSIRIK J. (2008). The bioscope corpus : biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, **9**(Suppl 11).
- ZHOU H., LI X., HUANG D., LI Z. & YANG Y. (2010). *Exploiting multi-features to detect hedges and their scope in biomedical texts*, In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning—Shared Task*. Association for Computational Linguistics.
- ÖZGÜR A. & RADEV D. R. (2009). *Detecting speculations and their scopes in scientific text*, In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing : Volume 3-Volume 3*. Association for Computational Linguistics.
- ØVRELID L., VELLDAL E. & OEPEN S. (2010). *Syntactic scope resolution in uncertainty analysis*, In *Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics.